THE GENOME ODYSSEY

MEDICAL MYSTERIES AND THE INCREDIBLE QUEST TO SOLVE THEM

Euan Angus Ashley, M.D., Ph.D.

CELADON
BOOKS
New York

Preface

I can't say for sure where my fascination with the genome began, but I'm pretty sure I know why it continues: unraveling the genome, our "code of life," is exhilarating. Day after day, we attempt to crack a code that is at once the essence of our humanity and also the destination for our pursuit of clues to help families afflicted with the most challenging genetic diseases. Life doesn't get more fulfilling than that.

None of this I could have imagined during a childhood growing up in the West of Scotland in the 1970s. That was around the time the first genome of a tiny organism was decoded. Since then, we've learned that your genome is a code linking you to every living organism on earth; a code incorporating the history of the human race; a code etched with your own family's history, dating back hundreds or even thousands of years. It is a code unique to you: no one alive and no one who has ever lived has the same one (not even your identical twin). Your code contains information about everything from your most likely height, weight, hair color, and eye color, to your predisposition, to some of many thousands of diseases. Your code can predict your future too: how you will live and how you may die.

All of this is inscribed by a molecule, deoxyribonucleic acid—DNA—

that has become so much a part of our lexicon that it has transfigured from molecule to metaphor. We say that a trait is "in the DNA" of a company or an institution despite the fact that neither of these is an organic being. We mean the trait is buried so deeply, bound so tightly, and wedded so irreversibly that it actually *is* the very fabric of its being. When we say of a fellow human that certain characteristics appear to be "in their DNA," we express the sense that these features are implicit, ingrained, and natural.

Paired in two strands, coiled into a right-handed double helix, DNA consists of just four simple building blocks arranged in patterns along its length. These nucleotides are the alphabet of life: A, T, G, and C—four letters forming the unique code that is locked in the vault of almost every cell in your body. Your genome is three billion letter pairs, six billion data points, two meters of molecule compacted into twenty-three pairs of chromosomes that, if laid end to end with the DNA from the thirty trillion cells in your body, could stretch to the moon and back thousands of times: part of the literal embodiment of what it means to be human.

The first human genome to have its letters spelled out (sequenced)—the Human Genome Project—was a ten-year, multicountry, multibillion-dollar effort to decode a mixture of DNA from ten people (actually, in the end, it was most of half a genome from more like two people, but more on that later). And while that first genome cost billions, it is, fewer than twenty years later, now possible to sequence a human genome for less than the cost of the cheap commuter bike I ride to work every day. This staggering reduction in cost has fueled a tsunami of scientific discovery and has given the medical profession an unparalleled opportunity to change lives for the better. Here is a tool for redefining disease, for solving medical mysteries, for providing hope to families who have suffered loss, and for protecting those who remain behind. This molecular microscope has allowed us to understand disease at a deeper level and to begin to personalize the practice of medicine.

The dramatic advances that you'll read about in this book—advances that have taken us from a theoretical understanding of DNA to the first human genome to a paradigm shift in medicine made possible by the sequencing of millions of genomes—came about because of revolutionary advances in our ability to read DNA and computer programs and human collaboration that have enabled us to understand it.

I think my younger self would have been surprised to hear I was dedicating my life to understanding the human genome even though, at some level, I was always going to be a doctor. Before the age of ten, I was the guy my school friends asked to wipe the blood off their grazed knees (my dad being the local doctor, I was the obvious choice, while my sister declared early as a veterinarian, so she got the dead birds). But I liked this role. I learned first aid around age twelve and became the resident expert among my peers on bodily functions. I bolstered this expertise by visiting my dad's medical office to inspect his tools and swivel on his doctor's chair. I was fascinated with his doctor's bag. It was black and box-shaped, came up to my knees, and appeared to me like a miniature hospital—a treasure trove of little drawers that opened to reveal needles, sutures, and intriguing metal implements, the utility of which I could barely imagine. I remember my dad once stitched a friend's forehead back together in our kitchen. That was extremely cool. Sometimes, he took me on his house calls. I remember one Christmas Day when he spent hours securing oxygen for one of his patients with lung disease to keep him out of the hospital. My mom was a midwife, and she would take me on her calls sometimes too. From both, I learned dedication, compassion, and that practicing medicine was not a job, nor even "just" a profession. It was a way of life, a state of being. I felt drawn to this on a visceral level: looking after people seemed like the thing I was put on this planet to do.

And yet, I was also definitively a technology geek. I remember

winning $150 when I was about ten and being forced into Scottish frugality, "investing" the money in a savings account rather than buying a pair of spectacles with built-in windscreen wipers. My dad and I bonded over solving the Rubik's Cube when I was about twelve (I think my record was thirty seconds, but the rose tint is obscuring the details now). I do clearly remember that summer when I was supposed to be outside enjoying the Scottish "weather," I was instead to be found inside more focused on writing a computer program to calculate the payroll taxes for my dad's medical practice. Of course, they changed the tax code the next year, teaching me an important lesson about, well, governments and taxes. I wasn't destined for accountancy anyway (my brother is the one to trust with your money), but I did imagine great wealth for myself from the profits of a horse racing game I wrote for the Sinclair ZX Spectrum home computer when I was fourteen. You made your horse go faster by alternately pressing the rubber keys as fast as you could. It didn't make me rich, but my nerd friends thought it was awesome. Or so they said.

At school, that teenage geek was drawn not only to physics and mathematics but also to language and music. My biology teacher, who informed my parents on one occasion that I was a buffoon, unwittingly ignited in me, at age sixteen, a fire for genetics when he gave me a copy of Richard Dawkins's *The Selfish Gene.* (The same teacher also suggested, with only marginal humor, that perhaps my experimental apparatus had fallen on the floor before I drew it in my scientific notebook. In his defense, art was never my strong suit.) My new pastime became reading popular science books—Dawkins, Gould, Lewontin, Sacks, Diamond, Pinker, and, of course, Darwin. I carried this passion with me into medical school. One of our early physiology classes sealed my fate. In groups of four, we sat in a lab at low benches on uncomfortable wooden stools, waiting anxiously to receive a recently excised still-beating rabbit heart. We poked it, felt its squidginess, picked it up, marveled at its

spontaneous beating, and hung it on a needle by the aorta, the blood vessel coming out of the top. Our job was to feed it, nurture it, and keep it alive as long as possible. I stared at this heart for hours. I was transfixed by its beauty—an elegant, exquisitely synchronized, biological machine. This wonder of evolution had a magical quality. It had rhythm and through its sounds produced music. I was hooked.

And so I found my medical calling. The specialty of cardiology had a vitality that was excitingly present: delivering electric shocks to patients whose hearts had stopped seemed to be an impactful way to spend one's day. And besides, many of my family suffered heart disease. Perhaps I could be useful to them one day.

Many years later, by virtue of a good dose of luck, I have found myself able to combine these passions as a practicing cardiologist and founding director of the Center for Inherited Cardiovascular Disease at Stanford University in California. The geek in me enjoys running a laboratory focused on using big computers and laboratory experiments to understand biology, and I have been fortunate, with talented colleagues, to cofound several biotechnology companies. As well as start-ups, I have been lucky enough to work with companies like Google, Apple, Amazon, and Intel, as well as the governments of several countries, advising them on the genome's place in medicine. I feel privileged to be alive at a time of technological transformation in genetics and health.

Today, you can sequence your genome for a few hundred dollars. Some years ago, I decided to devote myself to the cause of understanding the genome. It seemed to me that if we could parse its meaning, then perhaps we could start to decode our genetic future—to predict and prevent disease before it even began. I wondered if the quest might even inform, in some way, what it really means to be human.

This book tells of the first few years of adventures following that decision. Through its pages, I will take you on a journey into the science and medicine of the human genome. I will tell you stories of patients

whose care has been transformed by knowledge of their genome. I will introduce you to scientific teams I have led and admired and trace a path from genome data to medical action.

The first few chapters ("The Early Genomes") introduce you to our team and our efforts to medically decode some of the first genomes to be sequenced. I describe wandering into a Stanford colleague's office one day in 2009 to find him perusing his own genome and how that revelatory moment led us to try to bring every then known human genetic observation to bear on his genome to better understand his risks of disease. I tell of his cousin's son, who died suddenly in his teens, and how we used genetic sequencing of his postmortem heart tissue to try to find answers for his family. I tell of the origins of the human reference sequence in Buffalo, New York, and of the high school student who brought the genome sequences of her entire family to school for a science project. I describe how we extended these efforts into our Stanford clinics and tell how we started a company to broaden the impact beyond those walls.

In the second part of the book ("Disease Detectives"), I try to convince you that genomic medicine is very much like detective work. We start with the scene of a "crime" and carefully observe clues and document evidence using the tools of traditional medicine as it has been practiced for thousands of years: observe, examine, document, analyze. To this, we add our newest tool, reading the genome. I tell stories of patients with undiagnosed disease who have spent years looking for answers and find them through the wonder of genome science. I describe a national network of disease detectives whose mission is to systematically end these diagnostic "odysseys."

In the third section of the book ("Affairs of the Heart"), I describe some of my most treasured cardiac patients. I talk about a budding Broadway star for whom a play of genetic chance left her facing the consequences of an enlarged heart and the prospect of a heart transplant. I describe how we tried to break genome speed records to diagnose and

treat a newborn baby girl whose heart stopped five times on the first day of her young life. I relate how we found a young man with a big smile and multiple tumors growing inside his heart and how we traced the cause to a missing piece of his genome. I tell how we discovered a baby girl born with not just one but two different genomes. In this section, I also introduce you to the history of our understanding of heart disease and sudden death, the conditions that predispose to it, and the colorful characters whose work provided the insights we now use to treat it.

In the fourth and final part of the book ("Precisely Accurate Medicine"), I look ahead to the future. I discuss how examining superhumans—humans protected from disease by their genomes—can help make the rest of us just a little more "super." I describe several of the exciting new efforts to glean insight from sequencing millions of humans, including President Obama's Precision Medicine Initiative and the United Kingdom's Biobank. I talk about advances in curing and treating genetic disease, including genetic therapy, as well as methods to develop traditional drugs that are turbocharged by genomic insights.

No single book should attempt to cover all the ground, and so there are some important areas of medicine you will not read much about in this book. We do not dive deeply, for example, into cancer. As well as the fact that others have written eloquently on this subject, little of the amazing work in genetic testing of cancer patients and their tumors has involved sequencing the whole genome (most cancer tests involve deeply sequencing a subset of the genome). So the topic of cancer I have left to another day. I have also not included discussion about genetic tests during pregnancy. Our ability to pick up fragments of a baby's DNA in the maternal bloodstream has dramatically changed our approach to such testing, but these fragments are not routinely assembled into a whole genome, so I will hold on to those stories for now. Finally, there are many critical advances and compelling stories from my colleagues all around the world that I have not been able to tell in this book. I try to do justice to the stories of our own patients and our own

contributions in trying to help them, but there is so much work from others without whose endeavor ours would not have been possible.

The heroes of this book are my patients and their families. Together, they share a burden of inheritance: the little kids who smile and laugh or cry through their clinic appointments; the courageous teens coping with new diagnoses and living out unanticipated realities; the parents who get up every day to tend to the needs of their differently abled child. They travel from one doctor to the next, living every day the worry of a future unknown. They are the reason I get up every morning. Our patients and their families never escape what the genome brought to their lives, and we can never tire in our pursuit of understanding their genomes more fully, finding ways to treat their diseases more effectively, and, when our efforts fail, hugging them more tightly and assuring them that we will not rest until we come together upon a brighter day.

Part I

THE EARLY GENOMES

1

Patient Zero

Today we are learning the language in which God created life.

—UNITED STATES PRESIDENT BILL CLINTON

We share 51% of our genes with yeast and 98% with chimpanzees —it is not genetics that makes us human.

—DR. TOM SHAKESPEARE, UNIVERSITY OF NEWCASTLE

It was obvious to Lynn Bellomi that something was very wrong. It was August of 2011, and Lynn, from the city of Arroyo Grande on the scenic California coast, had just given birth to a beautiful boy named Parker. At first, everything seemed normal, but after several weeks, she became suspicious. Parker was still having trouble with things most babies learn to do fairly quickly, like feeding and sleeping. He was only sleeping a few hours a night. He cried a lot. By March 2012, when he was six months old, he was missing milestones—not showing curiosity about objects around him and not rolling over, never mind sitting upright. He was referred to a developmental specialist, then to an eye doctor, then to a brain doctor, then to a geneticist. To make matters worse, by nine months old, Parker appeared to be having regular seizures. He underwent many scans and dozens of tests, including painful blood draws. No one could figure it out. "Constant appointments, constant driving," his mom recalled. "And it felt like we were doing all these things without a purpose." Months turned into years.

In 2016, when we first met Lynn and five-year-old Parker, they had been referred to our Center for Undiagnosed Diseases at Stanford, part of a national network of doctor detectives whose aim is to solve the most challenging cases in medicine. Much of the time, success comes from analyzing a family's genomes, those DNA instructions that are the recipe book for all our cells and systems. So, on June 28, 2016, we drew blood from Parker so we could derive DNA from his white blood cells and spell out every letter in his genome. We also did this for his mom and dad.

Three months later, on October 4, genetic counselors Chloe Reuter and Elli Brimble called Lynn to say we had found a genetic change in Parker that did not appear to be inherited from either her or Parker's dad. It was a brand-new genetic mutation that arose in Parker, and it appeared to disrupt a gene called *FOXG1.* Other patients with damaging variants in this same gene had health problems that were remarkably similar to Parker's. This had to be the answer. For the first time since she had detected a problem with Parker's development, five years before, Lynn understood the size and shape of the enemy. She instantly gained a support group of families suffering with *FOXG1* syndrome around the world (650 parents on the *FOXG1* Facebook group, at last count). More than that, finally understanding the cause of Parker's disease allowed us to refer him to a movement disorder specialist who immediately changed his medications in a way that dramatically reduced his symptoms. "He still has some seizures, but much less frequently now," his mom recently told me. "He still has to go regularly to the doctor, but otherwise, he's a very happy guy."

Parker and his parents can now look forward to a world of new possibilities: joining with doctors, scientists, and hundreds of families from around the world to attack this disease from every angle, share experiences, disseminate insights, and hopefully, one day, find a cure. That future would have looked very different if it weren't for advances in our understanding of the genome—discoveries made, over the last

few decades, by scientists whose work has had a profound impact on the way we detect and treat human disease. To explore those breakthroughs, let's start by going back to 2009.

It was a pretty ordinary day. I had completed my morning meetings, and instead of lunch, I was heading to the office of a future friend, a Stanford physics professor and bioengineer named Stephen Quake. Steve was well known for his pioneering work in the field of microfluidics. He invented tiny biological circuit boards with switches, kind of like railroad "points," to direct cells or molecules to a specific micro-destination for analysis. Steve and I were meeting to plan an afternoon symposium for the genetics faculty at Stanford. His office is in a building at Stanford named for James H. Clark, an electrical engineer and founder of Silicon Valley companies like Silicon Graphics and Netscape. Designed by the famous British architect Norman Foster, the Clark building is shaped like a kidney with swooping red lines and lots of glass. At night, it is brightly illuminated and looks for all the world like an alien spacecraft that landed in the middle of campus. In a way, it kinda did. The building's purpose was to gestate a new specialty—bioengineering—the love child of a dalliance between biology and engineering. Situated on campus right in between the schools of medicine and engineering, against a California landscape of blue sky, sunshine, and palm trees, it is a stone's throw from the Stanford hospitals. Through its windows, as you walk by, you see brightly lit rows of worktops harboring the trade tools of engineering right next to the wet benches of molecular biology—robots interbreeding with pipettes. And during the day, after navigating a curious and gratuitously complex room numbering scheme, if you're lucky, you find Steve's office on the third floor.

Steve is the archetypal physics professor—Stanford and Oxford trained, a brilliant iconoclast. The breadth and diversity of his intellect emanates from a brain enveloped by tufts of professorial hair that,

in a former era, would be grown wildly to match his imagination. In fact, Steve's office is set up very much as I imagine his brain to be—mountains of chaotically "organized" scientific papers are piled up on every side and in every corner. He sits hunched in the middle, pecking at a keyboard, the creative energy source powering everything around. Amid a campus of overachievers, Steve stands out. I had gone that day to talk about a seminar we were running to bring together human geneticists across campus. But we never really got to that.

"Come look at this," he said. I found a place to sit amid the piles of journals, and he beckoned me over to look at the screen. It was not obvious at first what he was pointing at. There was a web browser open and a table filled the screen with the word *Trait-o-matic* at the top. It was one of those bare-bones spreadsheets with no formatting that was found on early websites—not pretty, but it was not the aesthetic that drew me in; it was the content. There were lots of columns of data. Gene names, gene symbols, As, Ts, Gs, and Cs, the building blocks of the genome.

"So what is that?" I asked.

His answer was to mark a pivotal moment for both of us. Delivered matter-of-factly and with more than a hint of his trademark understatement, it was as low-key as it was utterly revolutionary:

"It's my genome."

To put this into context, this was early 2009, and you could count on the fingers of one hand the number of people in the entire world whose genomes had been sequenced. Each one had been marked by an order of magnitude or more in reduction of cost. The Human Genome Project had been funded for $3 billion by the Department of Energy and the National Institutes of Health (NIH). And while each subsequent effort saw steep declines in cost, the price tags were still staggering. Craig Venter, the renegade entrepreneur who had taken on the public

genome project in a race to be the first to sequence a human genome, sequenced his own genome at a cost of around $100 million. An anonymous Han Chinese man had been sequenced in 2008 for around $2 million. And James Watson, who shared the Nobel Prize for work with Francis Crick and Maurice Wilkins and who, together with Rosalind Franklin, elucidated the structure of DNA, had his genome sequenced by a group at Baylor College of Medicine in early 2008 for the comparatively modest sum of only $1 million. Each of these projects involved hundreds of scientists and thousands of hours of time, as well as no small amount of blood, sweat, and tears. Then in 2009, Steve sequenced his own genome, in his lab, using a technology he invented himself, with postdoctoral scholar Norma Neff and Ph.D. student Dmitry Pushkarev, for just $40,000. In one week.

I was familiar with sequencing both in my research lab and in my clinic. We would send blood from our patients for DNA sequencing as a medical genetic test to try to find the cause of their inherited heart disease. Those tests would spell out the ATGC letters of the five to ten genes we knew could cause their heart condition to try to find the (usually) one-letter change that was the culprit. At that time, the cost for sequencing these five to ten genes was around $5,000 and the results took two to four months to come back. The test would provide an answer only about a third of the time, given that we were still in the early days of matching genes to diseases. So that was my context. To imagine we might have access to a whole genome—not five, not five hundred, not five thousand, but all *twenty thousand* genes as well as the other 98 percent of the genome that falls in between the genes . . . well, that was simply mind-blowing.

A few of us had, at that time, started to wonder, with the steep decline in the cost of genome sequencing, if one day patients might walk into our offices, figuratively or literally, "clutching their genomes." In Silicon Valley, we like to compare everything to computers, but the parallel between the rapid decline in the cost of sequencing and the rapid

decline in the cost of computing power was an enticing metaphor for many even beyond California's Bay Area. It became common among scientists to compare the drop in the cost of sequencing to Moore's law. Gordon Moore was a native of the Bay Area—a physicist who, along with Robert "Bob" Noyce, made fundamental contributions to the development of the integrated circuit, not least by starting Intel, one of the foundational semiconductor companies of Silicon Valley. In an article in 1965, referring to the rapid pace of technological advance, Gordon Moore observed that the rate at which components could be added to integrated circuits was doubling approximately every year, meaning the price of computing power was halving in this same period. He later decided two years was more realistic, but regardless, this "law" became synonymous with rapid technological advancement. It became common to show that the price of sequencing was dropping at a similarly spectacular rate, at least up until 2008, when the precipitous rate of decline in the cost of sequencing left Moore's law in the dirt. The National Human Genome Research Institute famously illustrated this by releasing a graph with a steep cliff-like falloff. I liked this graph and, like many genome researchers, would show it in my presentations. However, I soon found a more concrete, visceral way to put the price drop in perspective. My commute, at the time, took me past the Ferrari-Maserati dealership near Atherton—billionaire territory in the heart of Silicon Valley. I would often cast a sideways glance at those cars as I waited in traffic. One day, I was sitting at the stoplight doing random math in my head, as one does, and realized that if the Ferrari in the window had dropped in price as much as human sequencing had dropped in price in the eight years since the Human Genome Project's draft sequence was released, instead of $350,000 it would cost less than forty cents. A forty-cent Ferrari! A millionfold reduction in price. That seemed unprecedented. So I added that image to my slideshow. Sometimes, people tell me it's all they remember.

Admittedly, in 2009, with the cost of Steve's genome at $40,000, the

notion that patients would start bringing their genomes to clinic still seemed like a preposterously futuristic scenario, about as likely as my owning one of those Ferraris. But futurism is a potent driver of creative ideation. Shouldn't we start preparing for that day? Yes, there would be computational challenges and huge gaps in knowledge. But if we could meaningfully decode the genome, not just sequence it; if we could not just read the book but actually *understand* it; if we could turn data into knowledge, then put that to work for patients? Whoa.

So there I was in Steve's office, and he was asking me about various genes and pointing on the screen to the places where his own DNA letter was different from the one in the reference sequence (we will discuss the reference sequence and where it came from in chapter 6). "Do you see anything you recognize?" he asked. I scanned the names and noticed a gene I knew really quite well: cardiac myosin-binding protein C. This gene encodes for a protein that is an important part of the molecular motor of the heart. Its true function eluded scientists for years, but we now know that variants in this gene are the most common cause of the inherited heart disease hypertrophic cardiomyopathy—a disease associated with heart failure and sudden death. And here was Steve pointing to a variant in his genome in that very gene. There was a chance such a variant could be life-threatening. So naturally, being a cardiologist, I started asking him about his medical history. Do you have any medical conditions? Any symptoms? Chest pain? Shortness of breath? Palpitations? Instead of the scientist who walked into a colleague's office, I was now the physician talking to a patient: a very different kind of investigator, probing a very personal kind of truth. To my relief, Steve had not experienced any such symptoms and had no known medical condition.

So I moved to his family history. *Family history* means such different things to different doctors. To some, necessity requires it be a checkbox question: "Nothing in the family?" Move on. But to a geneticist or diagnostician of rare disease, the family history is a treasure trove,

a box of clues, to be scoured through, picked apart, examined, and deconstructed. This brand of diagnostician treats the family history like Sherlock Holmes treats the crime scene: examining it in minute detail, from every angle, actively interrogating, then reflecting. Yet few of us really know our family medical history well. Try it right now for yourself. Make a list of diseases that run in your family, and then try to match the names of the relatives who suffered from each disease along with how old they were when they were first diagnosed. Not so easy. I asked Steve the question, and like most patients, he answered quickly, "No, no family history of disease." And then, as if accessing some dusty file at the far end of the cabinet, "But wait, my dad has some heart thing, a rhythm problem . . . ventricular . . ."

"Tachycardia?" I offered, not expecting that to be right but, rather, reacting instinctively with the worst possible scenario (it's a doctor thing). Ventricular tachycardia is an abnormal heart rhythm that can occur, for example, in hypertrophic cardiomyopathy.

"Yeah, that sounds right."

Well, now, my curiosity was tinged with concern. In ventricular tachycardia, the normal coordinated activity of the top and bottom chambers of the heart is replaced by a rapid, dangerously uncoordinated rhythm that can be ineffective at pumping blood. Low blood flow to the brain means no consciousness and a rapid end to life. It is a rhythm that strikes fear into the heart of most doctors because it is almost always a medical emergency. Doctors *run* when they are called to patients in ventricular tachycardia. The name itself seems to lay down a staccato rhythm that evokes the broad, unruly electrical signal seen on the hospital monitors. It is a rhythm that screams, "Act now!" Fast, feared, and sometimes fatal.

So to recap, I had wandered into this meeting about organizing a genetics seminar, and here was my new friend Steve, the world-famous scientist, telling me that his dad possibly had reasons to suffer from ventricular tachycardia, a condition associated with sudden death. And

here was I, a cardiologist specializing in inherited cardiac diseases causing sudden death, staring at his genome at a particular variant in a particular gene known to be associated with hypertrophic cardiomyopathy, an inherited cause of sudden death. "So has anyone in your family ever died suddenly?" I asked. This question is arguably our most powerful tool. Such questions and their follow-up are, to a physician, as surgical tools are to a surgeon. Each surgeon has their favorite tools, some even personally crafted. They feel right in the hand. The balance is just so. The surgeon knows how her favorite tool responds, how it cuts. She understands innately the tissue response. When wielded in the right way, questions like this are the diagnostician's scalpel.

"Well, actually . . . my cousin's son recently died suddenly, and no one knew the cause."

Boom.

There it was: a family history of sudden unexplained death. The reddest of red flags, unfurled in front of me, waving in my face. Trying to sound casual, as I less-than-casually performed the mental math to calculate Steve's likelihood of sharing a genetic condition with his cousin's son, I said, "Oh, really, what sort of age was he?"

"Oh, he was only nineteen, a black belt in karate, I think, and never had a sick day in his life."

He had my attention now. The most common causes of sudden death in the young are inherited cardiac diseases like cardiomyopathy, like hypertrophic cardiomyopathy. And as I invited Steve to come to my clinic so we could check his heart, it was in that moment he became not just my colleague and my friend but my patient. And about a nanosecond after that, as my mind raced to work out how to put this together—how fast and what favors I would need to pull to screen his heart without delay—I realized that he was also about to be the first patient in the world to walk into a doctor's office for a checkup with his genome.

His complete genome.

And the doctor was me.

I headed back to my office, my mind racing with the possibilities and the impossibilities. How would one go about analyzing a genome, anyway? At the time, the idea of interpreting a whole human genome seemed as premature as it was preposterous. The handful of genomes released publicly at that point had undergone analyses that were mostly represented in statistics—this many single-letter variants were found, for example. The team at Baylor had gone further and looked at variants in medically relevant genes in Jim Watson's genome. But to think about scaling a medical approach to a whole genome, including every variant in every gene, was just not something anyone we knew had a viable solution for yet.

So I found one of my cardiology trainees, now a long-term collaborator and friend, an extraordinarily talented clinician-scientist named Matthew Wheeler. Matt is originally from upstate New York and trained in Chicago before coming to Stanford. He is tall, broad enough to power a rowing boat at speed, and nimble enough to look a whole lot better than I do skiing down a mountain. Our meeting was in fact orchestrated by our wives at a crew party for their rowing club, and we found shared passions in cardiology, genetics, sports, and inherited cardiovascular disease. On that day, we spoke about an ambition to build a Center for Inherited Cardiovascular Disease. This day, five years later, when we met in my office (which later became his office), I told him about Steve, his genome, his family history, and the idea that had formed in my head since walking back from my meeting: the idea, well, of clinically analyzing a whole human genome, every position, every gene, every variant. His response was typically understated, delivered deadpan, almost sotto voce, and foretold of the adventure upon which we were about to embark:

"Glad to see ambition hasn't left the building."

. . .

The human genome lives inside almost every cell in the body. I say "almost" every cell because certain cells, like red blood cells, for example, lose their nucleus as they mature—all the more room to pack in oxygen. The genome is housed in the cell's "inner vault," the nucleus, although some genes also live in the "power units" of the cell, the mitochondria. As mentioned earlier, the genome is made up of extremely long molecules of DNA. The individual chains of DNA are long strings of molecules called *nucleotides,* special sugars with one of four bases attached. The bases are adenine, thymine, guanine, and cytosine. The initial letter of each of these bases—ATGC—makes up the genetic code, six billion letters long. The DNA molecules that make up the genome are so long that if you were to stretch out the DNA from just one cell, it would be two meters long. That DNA needs to be compacted so it can fit into the nucleus. To achieve this, it is wrapped around proteins called *histones* and packaged into a compact structure called *chromatin* that makes up individual chromosomes. Each normal human genome has twenty-three pairs of these chromosomes: twenty-two regular ones and one pair of sex chromosomes, combinations of X and Y (females have two X chromosomes, males have one X and one Y). Some diseases emerge from duplications of whole chromosomes; for example, trisomy 21, a condition also known as Down syndrome, occurs if you have three copies of chromosome twenty-one. So to recap, the genome is a recipe book contained inside almost every cell in your body. It's six billion letters, all of them A, T, G, or C, and compacted into chromosomes of which most people have twenty-three pairs.

This recipe book contains ingredients and instructions for what to do with them. The ingredients are genes. These vary enormously in size: the smallest one is only eight letters, while the longest is 2,473,559 letters. Most genes represent the instructions for building a protein. To get there, the DNA is transcribed into a related molecule called *ribonucleic acid* (RNA) that carries the code as a message out of the nucleus for it to be translated, in groups of three letters at a time, into amino acids, the

building blocks of proteins that do the work of the cell. Proteins can be structural, to hold the cell together, or motor proteins, to move themselves or other things around, or enzymes that can convert one molecule to another. Yet the twenty thousand or so genes that account for all these proteins make up only about 2 percent of the genome. What about the other 98 percent? It seems almost unfathomable now that, for many years, this part of the genome was referred to as "junk DNA," reflecting the fact that nobody really knew what it was for. Our naïveté in assuming nature had no use for the vast majority of our genome embarrasses us further every year, as we learn more of the secrets of this "dark" genome. It turns out that this noncoding part of the genome is vital in determining whether genes are turned on or off. Also, about half of our genes have associated pseudogenes in this part of the genome—copies of the gene that are no longer functional (or so we used to think—now we know that pseudogenes can also regulate the function of other genes, especially their partner gene). Some of it sure *looks* a lot like junk; half the genome is made up of repeating sections of DNA that we still don't really understand. Finally, and perhaps craziest of all, almost 10 percent of our human genome is actually derived from viruses that embedded themselves long ago in our genome. Remember that the next time you have a cold.

Deciphering something as complex as the genome is something that seemed impossible for many years. In the 1970s, two approaches to reading DNA were proposed, but the one invented by Frederick Sanger came to dominate. Sanger was a British biochemist who, despite being one of only four people ever to win the Nobel Prize twice and mentoring two Ph.D. students who themselves won Nobel Prizes, used to describe himself as "just a chap who messed around in a lab." Sanger's approach, which dominated sequencing for decades and still plays a major role today, takes advantage of a molecular copying machine present in all our cells called *DNA polymerase*.

To understand Sanger sequencing, we're going to get a bit technical

for a minute. Think about starting with four tubes labeled A, T, G, or C. In each, we place our DNA copier, the DNA molecules we want it to copy, and the building blocks for making DNA (As, Ts, Gs, and Cs). Now, to each tube we add a special version of just one building block, the one corresponding to the label on the tube. The building block has a radioactive flag attached. When incorporated, this flag prevents the copier from lengthening that particular DNA molecule any further. Also, importantly, we add just a little of it compared to the regular building blocks. Now imagine as the copier in each tube starts, it grabs the building blocks it needs for the DNA molecule it is copying randomly from the mixture. Of course, it has a higher chance of incorporating regular building blocks than the special building blocks, because there are so many more regular ones. Eventually, however, by chance it grabs one with a flag. At that moment, the DNA copier is stopped in its tracks, and that molecule is flagged as radioactive. The copier moves on to make new copies elsewhere in the tube, and the cycle repeats. Eventually, there are four tubes each containing DNA copies of variable lengths. The A tube contains copies tagged with an A. The T tube contains DNA copies tagged with a T, and so on. To read the sequence, the DNA from each tube is taken out, and the molecules are spread out according to their lengths down a slab of gel using electrical charge. The radioactive elements can then be detected by exposing the gel to photographic film. The result is four tall, thin photographs, each of which appears like a ladder with lots of rungs missing. However, here's where the magic happens. If you line up the four photographs next to each other, you see that each rung is represented in only one of the photos. The ladder in which the rung appears corresponds to the letter for that position: A, T, G, or C.

If you didn't catch all of that, just bear with me. This laborious process was accelerated and commercialized with three major advances: 1) radioactivity was replaced with light-emitting molecules, 2) everything was run in one tube, and 3) molecules were separated much faster and

more efficiently based on their electrical charge. This technology, reading DNA copies each about five hundred letters long, was developed by the company Applied Biosystems and became the workhorse sequencing approach for the Human Genome Project.

The second genome, completed using this same technology, was finished around the same time as the Human Genome Project and belonged to Craig Venter, a scientist who had formed a company to sequence and attempt to patent human genes. He created a firestorm by challenging the public program to a race to the finish (in the end it was declared a draw). Venter's genome cost around $100 million to sequence (representing a staggering drop in price for our Ferrari from its original $350,000 to a mere $12,000).

Many such breakthroughs in biology come to parallel, if not the content, then certainly the language of science fiction. And that may, or may not, be the reason that so-called next-generation sequencing was born. *Star Trek*'s Jean-Luc Picard would have been proud. And of course, since *next* is a relative—not an absolute—term, it was perhaps inevitable that almost everything since Sanger sequencing has been referred to as *next generation* at one point. It is truly the gift that keeps on giving. And that gift is confusion. But the thing that all next-generation technologies have in common is their ability to expand the process of sequencing. Instead of focusing on the part of the genome you want to sequence, making many copies of just that part, then Sanger sequencing those, with next-generation sequencing, you take the whole genome, chop it up into small pieces around one hundred letters, then sequence all the pieces at the same time. This allows, effectively, a massive turbocharge to sequencing.

Such a technological advance took some time to finesse. It was to be seven years before another individual's genome was published. In 2007, the Nobel Prize winner James Watson's genome was sequenced using a technology from a company called 454 (founded by serial entrepreneur Jonathan Rothberg) by a team at Baylor College of Medicine led by

Australian geneticist Richard Gibbs. Roche bought the mysteriously named 454 technology in 2007 because of its ability to sequence very long pieces of DNA (initially four-hundred- to five-hundred-letter fragments; it was later updated to read fragments up to one thousand letters). According to Baylor's analysis, Watson's genome revealed a predisposition to cancer. He also famously redacted from the public disclosure of his genome the status of a gene variant predisposing him to Alzheimer's disease. His genome took two months to complete and cost $1 million. That Ferrari was just discounted to $116.

Various groups around the world published three more genomes (these were anonymous) in quick succession at the end of 2008 and beginning of 2009. All were sequenced using technology from a company called Illumina, the dominant force in sequencing for most of the last ten years. Importantly, these genomes started to represent more of the world's diversity: one individual was Han Chinese, one was Korean, and the other was West African. The last publication included some medically oriented annotation of the genome and even used an early version of the Trait-o-matic software that I first saw in Steve's office. Each took about six to eight weeks to complete and cost a few hundred thousand dollars: a trio of $50 Ferraris.

Steve's genome stood out for several reasons. For starters, he invented the technology used to sequence the genome and founded a company, Helicos, to market the instrument he invented—the cutely named HeliScope. The Helicos approach differed from the Sanger and Illumina techniques, because it sequenced single molecules of DNA. Fluorescently labeled DNA bases were flooded into a device called a flow cell where the short target stretches of DNA were anchored. As each base was incorporated into a new strand of DNA by DNA polymerase, the copier, a very sensitive camera, would take a picture, kind of like taking a photo of a tiny light bulb. Then, a wash step would chop off that light bulb, and another would be flowed in. Then another photo would be taken, and the cycle would repeat. But of course, each

photo wasn't just one light bulb. The camera could read a billion bulbs at once, meaning that enough data to cover a whole human genome could be generated *in one week,* at a cost of $40,000. Today, sir, your Ferrari will be assembled in just under one hour and discounted to $6.

As you might imagine, all these next-generation approaches output millions of short genome "words" that correspond to the small fragments of DNA that are fed into the sequencer. The words do not come out in any particular order, so to be understood, they need to be organized—put together kind of like a jigsaw. That is usually done via a computer program that scans the human reference sequence (the sequence created by the Human Genome Project) and locates the correct position for each new word. Such programs are standard now, but at the time, the software had to be written from scratch. That job fell to Dmitry Pushkarev, from Steve's lab, a tall, lean Russian graduate student with enviable stamina both in late-night coding and daytime adventure pursuits. Dmitry built some of the first programs to stitch together a genome and find the places where it varied from the human reference sequence. And it was in that data and those algorithms that our work began.